



# Phrase-Based Machine Translation based on Simulated Annealing

Caroline Lavecchia, David Langlois, Kamel Smaïli

## ► To cite this version:

Caroline Lavecchia, David Langlois, Kamel Smaïli. Phrase-Based Machine Translation based on Simulated Annealing. Sixth international conference on Language Resources and Evaluation - LREC 2008, May 2008, Marrakech, Morocco. inria-00285277

**HAL Id: inria-00285277**

**<https://inria.hal.science/inria-00285277>**

Submitted on 5 Jun 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Phrase-Based Machine Translation based on Simulated Annealing

Caroline Lavecchia, David Langlois and Kamel Smaïli

LORIA - Speech Group  
Campus scientifique, BP 239, 54506 Vandoeuvre lès Nancy Cedex, France  
lavecchi, langlois, smaili@loria.fr

## Abstract

In this paper, we propose a new phrase-based translation model based on inter-lingual triggers. The originality of our method is double. First we identify common source phrases. Then we use inter-lingual triggers in order to retrieve their translations. Furthermore, we consider the way of extracting phrase translations as an optimization issue. For that we use simulated annealing algorithm to find out the best phrase translations among all those determined by inter-lingual triggers. The best phrases are those which improve the translation quality in terms of Bleu score. Tests are achieved on movie subtitle corpora. They show that our phrase-based machine translation (PBMT) system outperforms a state-of-the-art PBMT system by almost 7 points.

## 1. Introduction

Given a sentence in a source language, the goal of Machine Translation (MT) is to find out its translation in a target language. Different approaches exist to deal with this difficult challenge. Some approaches require *a priori* human knowledge in order to model both the source and target languages, and how to switch from one to another. The Systran MT system (Jean Senellart, 2001) is based on this approach and proposes a translation model depending on transfer rules. The statistical approach follows a completely different direction.

The statistical MT does not require any external knowledge. It uses only parallel corpora to model the translation process. Such corpora are aligned at word or sentence level in order to link both source and target languages. The translation issue is treated as an optimization problem. Translating a sentence from English into French involves finding the best French target sentence  $f^*$  which maximizes the probability of  $f$  given the English source sentence  $e$ . This translation model is based on the noisy channel model. The Bayes rule allows to formulate the probability  $P(f|e)$  as follows:

$$f^* = \operatorname{argmax}_f P(f|e) = \operatorname{argmax}_f P(e|f) * P(f) \quad (1)$$

Thus, the translation process consists of a language model  $P(f)$  and a translation model  $P(e|f)$ . Language model parameters are trained from a target corpus, whereas parameters of the translation model are determined from the parallel corpus. Then, a decoder provides the best target sentence given the source sentence and the table translation parameters.

First statistical MT systems were word-based (Brown and al., 1993). Obviously, the human translation is a very complex process which is not only word based. Following this fact, recent researches showed that the use of phrase translation instead of word translation leads to better MT system quality. Dealing with phrases allows an explicit modeling of lexical units and captures easily local reordering. For example, without use of phrases, the translation

of *Pomme de terre* gives *apple of earth* instead of *potatoe*. And in some situations, the use of phrases reduce the imprecision of reordering. For instance, without use of phrases, the translation of *Tour Eiffel* gives *Tower Eiffel* then reordering process may produce the correct English translation. By using phrases, we reduce the imprecision of translation and at least avoid some reordering problems. Probably, one of the most difficult issue is how to find out the best phrases in both source and target languages.

In order to retrieve phrases, several approaches have been proposed in the literature. Most of them require word-based alignments. For example, (Och et al., 1999) collected all phrase pairs that were consistent with the word alignment provided by Brown's models. Thus any contiguous source words must be the translation of any contiguous target words on the condition that words are aligned with each other. That means that retrieved phrases have not always linguistic motivation and could lead to noisy sequence of words.

In this paper, we propose an original idea based on inter-lingual triggers to build phrase translation without requiring word-based alignments. First we give an overview of inter-lingual triggers. Then we present the set up of our phrase-based machine translation system based on inter-lingual triggers. Finally, a description of the used corpora and the results are provided and discussed. We end with a conclusion which points out the strength of our method and gives some tracks about future work in our research group.

## 2. Machine Translation based on Inter-Lingual Triggers

We propose an original approach for SMT based on inter-lingual triggers. In the following, we present the notion of inter-lingual triggers and how to make good use of them in order to perform Machine Translation.

### 2.1. Review of inter-lingual triggers

Inter-lingual triggers are inspired by the concept of triggers used in statistical language modeling (Tillmann and Ney, 1997). A trigger is a set composed of a word and its best

correlated triggered words in terms of mutual information (MI). Trigger models are combined with  $n$ -gram models in order to enhance the probability of triggered words given a triggering word.

Since classical triggers allow to establish a triggering-triggered link between two events from the same language, we propose to determine correlations between words in a source language and words in a target language by using inter-lingual triggers. Therefore, an inter-lingual trigger is a set composed of a triggering source event and its best correlated triggered target events. We hope to find among the set of triggered target events, possible translations of the triggering source event. Inter-lingual triggers are determined on a parallel corpus according to the following formula:

$$MI(f, e) = P(f, e) * \log\left(\frac{P(f, e)}{P(f) * P(e)}\right) \quad (2)$$

where  $f$  (respectively  $e$ ) is a sequence of French (respectively English) words.  $MI(f, e)$  denote the mutual information assigned to  $e$  and  $f$  and  $P(e)$ ,  $P(f)$  and  $P(f, e)$  are defined as follows:

$$P(X) = \frac{N(X)}{|Corpus|} \quad P(f, e) = \frac{N(f, e)}{|Corpus|} \quad (3)$$

where  $N(X)$  is the number of sentences where  $X$  occurs,  $N(e, f)$  is the number of sentence pairs where  $e$  and  $f$  co-occur and  $|Corpus|$  is the number of sentence pairs in the training corpus.

For each French event  $f$ , we kept as inter-lingual triggers, the  $k$  English events  $e$  with higher MI values. In the following, an event is a word or a sequence of words and we differentiate two types of inter-lingual triggers:

- **1-To-1 triggers:** one French word triggers one English word
- **n-To-m triggers:** a sequence of  $n$  French words triggers a sequence of  $m$  English words with  $n, m \in \mathbb{N}$ .

Inter-lingual triggers have been used in (Kim and Khudanpur, 2004) to enrich resource deficient languages from those which are considered as potentially important. Our purpose is to use them in order to perform statistical machine translation. To achieve that, we employ inter-lingual triggers to build translation tables required in the decoding process. To do that, we assign to each inter-lingual trigger a probability calculated as follows:

$$\forall f, e_i \in Trig(f) \quad P(e_i|f) = \frac{MI(e_i, f)}{\sum_{e \in Trig(f)} MI(e, f)} \quad (4)$$

where  $Trig(f)$  is the set of  $k$  English events triggered by the French event  $f$ .

In a previous work, we developed a Word-based Translation system based on 1-To-1 triggers (see section 2.2. for more details). In this paper, we extend inter-lingual triggers to carry out phrase-based Machine Translation. In the following, we present our method to build phrase translation table based on Simulated Annealing.

## 2.2. Word-based Translation with Inter-Lingual Triggers

In (Lavecchia et al., 2007b), we built a word-based Machine Translation (WBMT) system based on 1-To-1 triggers. First, we constructed a word translation table using the 50 best triggers for each French word. Then, we used the Pharaoh decoder to translate an English corpus into French. We showed that the performance of our system is similar to the ones achieved by a system based on IBM model 2 (Brown and al., 1993), in terms of Bleu score (Papineni and al., 2001).

In the light of this supporting results, we decided to investigate phrase-based Machine Translation (PBMT) based on inter-lingual triggers. As we have seen before, most of state-of-the-art methods collect phrase translation from word-based alignments. Our goal is to train a PBMT system without calling upon word alignment. We would like to learn phrase pairs only by taking advantage of inter-lingual triggers.

## 2.3. Method for learning phrase translation

Most of methods which use phrases in MT require word-based alignments. For example, (Och et al., 1999) collected all phrase pairs that were consistent with the word alignment. In his method, any contiguous source words may be the translation of any contiguous target words on the condition that words are aligned with each other. That means phrases have no always linguistic motivation and retrieved translations could lead to noise.

We are convinced that if we succeed in identifying common phrases in the source part of the training corpus, inter-lingual triggers will allow to retrieve its translations in the target part. This would generate less noise. Since source phrases are selected beforehand, our method does not require any word alignment. In the next sections, we detail how to extract source phrases. Then, we propose to use inter-lingual triggers in order to find their potential translations in the target corpus. Finally, we present an adaptation of the Simulated Annealing algorithm in order to determine the best phrase translations among all those selected with inter-lingual triggers.

### 2.3.1. Phrase extraction

In the few last years we developed a statistical method to extract pertinent phrases (Zitouni et al., 2003) from large corpus. We use this method to rewrite source part of the training corpus in terms of phrases. To achieve that, an iterative process selects phrases by grouping words which have a high value of Mutual Information. Only the phrases which improve the perplexity are kept for the forthcoming steps. At the end of the process, we get a list of phrases and a source corpus rewritten in terms of phrases. With this source corpus expressed with pertinent phrases, we hope to find their potential phrase translations in the target corpus by using inter-lingual triggers.

### 2.3.2. Learning phrase translation

The source training corpus is henceforth rewritten in terms of phrases. Now, the question is how to find the potential translations of these source phrases in the target corpus. To

achieve this, we propose to use inter-lingual triggers. In the following, we assume that each source phrase of  $l$  words can be translated by a sequence of  $j$  target words where  $j \in [l - \Delta l, l + \Delta l]$ .

At this step, no word alignment is performed. For this reason, we associate with each source phrase  $(2 * \Delta l + 1)$  sets of its  $k$  best inter-lingual triggers. Thus, we allow a source phrase to be translated by different target sequences of variable sizes. Table 1 shows the potential translations of the source phrase *porter plainte*. In the following we

source phrase	n-To-m triggers		
	2-To-1	2-To-2	2-To-3
porter plainte	press	press charges	can press charges
	charges	can press	not press charges
	easy	not press	you can press

Table 1: Potential translations of the source phrase *porter plainte*

guess that for short phrases  $\Delta l$  is set to 1. Thus, for the cited example, we suppose that it could be translated by a sequence of at least one word and at most by a sequence of 3 words. For this reason, we associate it with its best 2-To-1, 2-To-2 and 2-To-3 inter-lingual triggers. In this example, we have selected 9 potential translations. Obviously, only *press charges* is a correct one. In the general case, we can have for each phrase of two words  $k$  potential translations. That is why we propose to select those which are pertinent and discard the noisy ones.

All source phrases and their sets of inter-lingual triggers constitute the set of n-To-m inter-lingual triggers. Now, the issue is how to select the best n-To-m inter-lingual triggers. In other words, what are the pertinent phrases and their translations. To answer this question, we first compute Bleu score on a development corpus by using our word-based system based on 1-To-1 inter-lingual triggers. This will constitute the baseline result. In a second step, we add randomly a subset of n-To-m triggers previously computed into the word-based system. With an adequate algorithm, we select the most relevant phrases, those which improve the Bleu score on a development corpus. The optimization algorithm we use is simulated annealing detailed in the next section.

An outline of retrieving the best phrase translations is given in Algorithm 1.

---

**Algorithm 1** Method for learning and selecting the best phrase translations

---

- 1: Extract phrases from the source corpus
  - 2: Determine n-To-m inter-lingual triggers which allow to associate each source phrase with the best target phrases of variable size
  - 3: Compute the baseline Bleu score by using our word-based system based on 1-To-1 inter-lingual triggers.
  - 4: Select an optimal subset of n-To-m inter-lingual triggers on an iterative process handled by Simulated Annealing algorithm
- 

### 2.3.3. Simulated Annealing tuning

Simulated Annealing (SA) algorithm is a technique applied to find an optimal solution to a combinatorial problem that becomes unmanageable using combinatorial methods. The SA approach allows to solve such combinatorial problem while dealing with the local optimum problem. The concept of SA is inspired from the physical annealing process of solids and is easily adaptable to solve large combinatorial optimization problems. In condensed matter physics, people are interested in obtaining low energy states of a solid. In other words, the issue is how to arrange the billions of particles in order to achieve a highly structured lattice with a low energy of the system.

Our aim is similar, in fact, the set of n-To-m triggers constitute a list of candidate phrase translations. We have to integrate a subset of this list in our MT system in order to increase the quality of translation. Naturally, it is unreasonable to try all possible combinations of translation. For this reason, we decided to use SA algorithm in order to select the ones which lead to the best performance. To achieve that, we start with a word-based MT system based on 1-To-1 inter-lingual triggers. Then we randomly add n-To-m triggers into the MT system until an optimal Bleu score is reached on the development corpus.

The entire algorithm is given below:

---

**Algorithm 2** Simulated Annealing algorithm

---

- 1: Start with a high temperature  $T$ .
  - 2: With a temperature  $T$  and until the equilibrium is reached do  
 From the current temperature  $T$  of the system and from the current state  $i$  which has an Energy  $E_i$ , perturb the system which makes it moving from state  $i$  to  $j$ . The energy of state  $j$  is  $E_j$ .  
 If  $E_j - E_i \geq 0$  then state  $j$  is accepted as the current state; Otherwise, state  $j$  is accepted with a probability  $random(P) < e^{(E_i - E_j)/T}$  with  $P \in [0 - 1]$
  - 3: Decrease the temperature and go to step 2 until the given low temperature is reached or until the energy stops increasing
- 

It is necessary to define all the parameters of the algorithm in order to adapt it to our issue:

**Initial temperature** The temperature acts as a control parameter. Several values have been tested in our experiments for the initial temperature.

**Initial configuration** The initial state is a word-based MT system based on 1-To-1 inter-lingual triggers.

**System perturbation** Agitate the system consists in randomly adding a subset of n-To-m inter-lingual triggers into the translation table of the current MT system.

**Equilibrium State** At each step of the SA algorithm, a whole decoding process is launched in order to evaluate the performance of the current MT system in terms of Bleu score. The equilibrium state is reached when the Bleu score stops increasing between two states.

**The schedule annealing** After each equilibrium state, the temperature has to be decreased carefully. For that, we choose a geometric series, which respects the progressive decreasing of the temperature.

**Energy computing** The energy to be maximized is expressed by the Bleu score on a development corpus.

**Stop criterion** The stop criterion of adding n-To-m inter-lingual triggers is reached when the Bleu score of the system converges.

At the end of the SA algorithm, only the n-To-m inter-lingual triggers which improve the performance of the initial word-based MT system are selected. In SA algorithm, skipping from one state to another guarantees to reach an optimal state in terms of the objective function. Consequently this algorithm increases necessarily the Bleu score. In the next section, we present used corpus and conducted experiments to train and test our PBMT System based on inter-lingual triggers.

### 3. Results

#### 3.1. Corpora

We present results on a subtitle parallel corpus built using Dynamic Time Wrapping algorithm (Lavecchia et al., 2007a). Subtitle corpora are very attractive due to the used spontaneous language which contains formal and informal words. We think that such corpus constitute a good challenge to go towards spontaneous speech translation system. Table 2 gives details about the used the parallel corpus. We use a train corpus to extract French phrases and to compute inter-lingual triggers (a study of few examples is given in section 3.2.). A development corpus is used to select the best phrase translations among all those determined by the set of inter-lingual triggers. Finally, we use a test corpus to validate our approach.

		French	English
Train	Sentences	27523	
	Words	191185	205785
	Singletons	7066	5400
	Vocabulary	14655	11718
Dev	Sentences	1959	
	Words	13598	14739
Test	Sentences	756	
	Words	5314	6262

Table 2: Quantitative description of the training corpus (Train), the development corpus (Dev), the test corpus (Test)

As shown in Table 2, more than 45% of the words in both French and English vocabularies occur only once in the training corpus. Furthermore, 14.5% (respectively 13.8%) of the English words in the development (respectively test) corpus are out of vocabulary (OOV). All these elements account for weak Bleu scores reported in section (3.).

In the following paragraphs, we present a study of few inter-lingual triggers.

#### 3.2. Study of some inter-lingual triggers

Inter-lingual triggers are selected on a parallel training corpus. In our framework, training step leads to significant inter-lingual triggers as shown in Table 3.

French	English	$MI \times 10^{-4}$
porter plainte	press charges	6.92
	charges	6.26
	press	5.29
allumer	light	4.65
	to turn on	3.46
	turn on	2.88
bonjour	hi	32.43
	hello	29.30
	good morning	19.55
calme toi	calm down	23.32
	calm	21.99
	down	13.85
petit déjeuner	breakfast	10.42
	to breakfast	3.13
	say breakfast	3.138

Table 3: Examples of English phrases triggered by French phrases

The first column presents French sequences of one or several words. Sequences that have more than one word are automatically picked up by the iterative process explained in section (2.3.1.). For each French sequence, the second column refers to the best correlated English sequences of one, two or three words in terms of MI. Finally the third column shows the MI value associated with each inter-lingual trigger. A qualitative analysis showed that our method leads to pertinent inter-lingual triggers. Thus, triggered sequences could often be considered as potential translation of the triggering French sequence. Furthermore, inter-lingual triggers allow to retrieve synonyms as it is shown for the French word *allumer* which can be translated by *light* or *turn on*. Note also that they take into account the fact that  $n$  French words are not necessarily translated into  $n$  English words. Thus, *bonjour* is associated with *good morning* or *petit déjeuner* with *breakfast*. Finally, when a French sequence is translated into several English words, inter-lingual triggers will prefer the whole English sequence rather than subparts of it. This case is illustrated by the example *porter plainte* which is translated by *press charges*, the sequence which gets the highest MI value.

In the following sections, we evaluate our MT systems based on inter-lingual triggers. To achieve that, we use inter-lingual triggers to build translation tables required by the decoder Pharaoh (Koehn, 2004) in order to translate an English source corpus into French. Bleu score allows us to evaluate the quality of the obtained French translation. In section (3.3.), we start with a Word-Based Machine Translation (WBMT) system based on 1-To-1 triggers. Then, we compare it with a state-of-the-art WBMT system based on IBM model 3. In section (3.4.), we design a Phrase-Based MT (PWMT) system based on n-To-m triggers and

Simulated Annealing. Finally, we compare it with a state-of-the-art PBMT system based on the phrase-based model proposed by Och in (Och, 2002).

### 3.3. Word-based Translation System

To build our WBMT system, we employ 1-To-1 triggers for which each French word is associated with a list of  $k$  English words. We hope to catch in this set of  $k$  English words potential translations of the French word. Several experiments showed that 10 is the optimal value for the parameter  $k$ . Then, we assign to each 1-To-1 trigger a probability calculated from  $MI$  as indicated in formula (4). This constitutes the word translation table required by Pharaoh.

System	tm	lm	d	w	Bleu
1-To-1 Triggers	0.6	0.3	0.3	0	<b>12.49</b>
IBM Model 3	0.8	0.6	0.6	0	12.39

Table 4: Evaluation of WBMT systems

Translation results in terms of Bleu on the development corpus are given in Table 4. The first line of the table reports performance of our WBMT system based on 1-To-1 triggers. The performance is compared to the one of a WBMT system based on IBM model 3 reported on the second line. For an optimal use of the decoder, the weights of the models involved in the decoding process are tuned on the development corpus<sup>1</sup> for both systems.

Results show that using 1-To-1 triggers leads to better translation quality. Indeed, the better performance of our system amounts to 12.49 in terms of Bleu score. While in an optimal use, the system based on IBM model 3 reaches only 12.39. Furthermore, this last model is trained in several iterations whereas training inter-lingual triggers needs only one iteration. In other words, with less time for training, our approach leads to better results than famous IBM models largely used in SMT.

Considering this very promising results for WBMT, we decided to make good use of inter-lingual triggers to process Phrase-based Machine Translation.

### 3.4. Phrase-based Translation System

#### 3.4.1. French extracted phrases

To build our Phrase-based Machine Translation System, we extracted from the French part of the training corpus, a set of 15860 phrases which are composed of two or three words.

Only 2.20% (respectively 3.03%) of the phrases extracted from the training corpus were in the development (respectively test) corpus.

<sup>1</sup> $tm$  (respectively  $lm$ ,  $d$ ) indicates the weight of the translation (respectively target language, distortion) model. The parameter  $w$  is for the word penalty. The target language model is a trigram model (Good-Turing smoothing)

#### 3.4.2. Candidate Phrase translations for SA algorithm

For each French unit (a word or a sequence) of  $l$  words, we select from the training corpus its  $10 * (l \pm \Delta l)$  best inter-lingual triggers. For practical reasons,  $\Delta l$  does not exceed 2. This means, for each potential English translation set among those containing  $l - \Delta l, l - \Delta l + 1, \dots, l + \Delta l$  words, we kept the best 10 units. All this inter-lingual triggers make the set of candidate phrase translations (called n-To-m triggers) required by SA algorithm.

#### 3.4.3. Results with Simulated Annealing

Different experiments have been conducted to optimize the parameters of the SA algorithm. In this section, we present the performance with the optimal set of parameters. We made several tests in order to determine the best value of the initial temperature.  $T = 10^{-4}$  seems to be a convenient initial temperature. The initial configuration consists of the word translation table obtained in section (3.3.) with 1-To-1 triggers. This configuration leads as we have shown, to an initial energy of 12.49 in terms of Bleu score. Then, at each step of the SA algorithm, we agitate the current configuration by adding randomly phrase translations from the set of n-To-m triggers selected earlier on the training corpus. Conducted experiments showed that performance are optimal when we added randomly 10 potential translations of 10 French words or phrases.

The improvement of the Bleu score obtained on the development corpus through SA algorithm is shown in Figure 1. At the end of the SA process, our phrase-based MT system

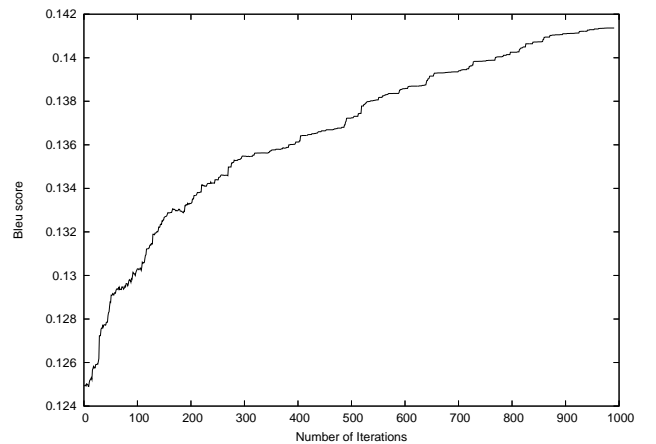


Figure 1: Improvement of the Bleu score on a development corpus through the SA algorithm

fulfilled a Bleu score of 14.14. In other words, by adding pertinent phrase translations, we achieved an improvement of more than 1.6 point in terms of Bleu compared to our word-based MT system.

In the next section, we compare our PBMT system based on inter-lingual triggers with a state-of-the-art PBMT in order to evaluate our approach.

### 3.5. Comparison with a state-of-the-art system

In order to validate our approach, we compare the performances of our PBMT system based on inter-lingual triggers with a state-of-the-art PBMT system (reference system). The phrase translation table of the reference system is acquired from a word-aligned parallel corpus by extracting all phrase-pairs that are consistent with the word alignment (Och, 2002). Table 5 illustrates the performances of

	inter-lingual triggers		state of the art	
	1-To-1	n-To-m	IBM3	reference
Dev	12.49	14.14	12.39	7.02
Test	13.63	10.77	14.00	6.57

Table 5: System Evaluation in terms of Bleu score on the development (Dev) and the test (Test) corpora

the different systems on both development and test corpora. The two first column denote the Bleu score achieved by our WBMT (1-To-1) and PBMT (n-To-m) systems based on inter-lingual triggers. Whereas the two last columns correspond to the performance of the state-of-the-art WBMT system based on IBM model 3 (IBM3) and the reference PBMT (reference).

On the development corpus, as seen before, the use of pertinent n-To-m triggers improved the results achieved by 1-To-1 triggers by almost 13.21%. For the state-of-the-art methods, the use of phrases decreased the performance by 43.3% compared to the word-based method. Despite the few amount of training data, these first results show that the SA algorithm allowed to take off the phrase translations with no statistical significance. Overall phrase translations determined by n-To-m triggers, it selected only a subset which leads to an optimal translation quality. Furthermore, both WBMT and PBMT systems based on inter-lingual triggers led to better performances than the corresponding state-of-the-art systems.

Unfortunately, the lead of n-To-m triggers on 1-To-1 triggers is not corroborated on the test corpus. Indeed, the use of phrase translations decreased the Bleu score by 21%. Over-fitting due to poor amount of data and used corpora can explain this under-achievement. Recall that the used corpora are subtitles of 36 different movies. Each movie is divided into three parts: one for the training corpus, one for the development corpus and one for the test corpus. And we have chosen movies without paying attention of the different cine styles. Consequently, talks and expressions are very disparate within each corpus. For example, the talks in a thriller will not be the same as in a comedy. By adding the poor amount of data, pertinent phrases chosen on the development corpus were not necessarily pertinent on the test corpus. Conversely, phrase translations unselected by the SA algorithm would maybe allow to improve performance on the test corpus. A good way to figure out this problem would be to classify movies according to their style (thriller, comedy, war movie, romantic comedy ...) and to set one PBMT system based on inter-lingual triggers by style. Hence, even if data are sparse, learned phrase translations would be proper to each movie style. And phrase

translations selected during SA process on the development corpus would be the best ones on the test corpus too.

Anyway, the impact of over-fitting are more important on the state-of-the-art systems. Indeed, on the development corpus, their performance decreases by 43.3% from WBMT to PBMT. Whereas for systems based on inter-lingual triggers, Bleu score increases by 13.9%. In the same manner, on the test corpus, state-of-the-art PBMT system decreases the Bleu score of the state-of-the-art WBMT system by 53.07%. While n-To-m triggers PBMT system lowers the performance of 1-To-1 triggers by only 21%.

## 4. Conclusion

In this paper, we presented our phrase-based Machine Translation system based on inter-lingual triggers. The latter allow to associate a triggering source phrase with its best triggered target phrases in terms of mutual information. We noticed that a triggered target phrase may often be assimilated to a potential translation of the source phrase. Thus, we decided to use inter-lingual triggers in order to set up our PBMT system.

Most phrase-based translation models require word alignment on the parallel training corpus. Phrase translations extracted from this alignment are not always linguistically motivated and thus are not pertinent. In order to extract more relevant phrase translations and therefore improve the translation quality, we proposed an original method that does not need any word alignment. First we identified common source phrases by an iterative process. Then, we retrieved their potential translations by using inter-lingual triggers. And finally, we used simulated annealing algorithm to select the best phrase translations among all those determined by inter-lingual triggers.

We trained and tested our PBMT system on a subtitle parallel corpus built using Dynamic Time Wrapping algorithm. This corpus constitutes a good challenge to go towards spontaneous speech translation system. Once phrase translation table were induced by inter-lingual triggers and simulated annealing algorithm, we used the decoder Pharaoh to translate text from English into French. We evaluated the translation quality with the Bleu metric. Results showed that our approach led to better translation quality compared to a state-of-the-art phrase-based approach that required word alignment. Indeed, our system based on inter-lingual triggers outperformed a state-of-the-art system by 7 points on a development corpus and by 4 points on a test corpus. Conducted experiments confirmed that phrase translations learned from word alignment can cause noise in the translation process. Identifying common source phrases and selecting their potential translations with inter-lingual triggers and SA algorithm allows to confine noise even on sparse data.

Our results are very encouraging and efforts are done in order to improve our model. The idea of using inter-lingual triggers seems to be very important. For the moment, we focus on word surface forms. However, considering inter-lingual triggers on syntactic features in order to integrate linguistic knowledge in the translation process may improve drastically the translation quality.

## 5. Acknowledgments

This work is supported by EADS (European Aeronautic Defense and Space Company) foundation for the Speech-To-Speech translation Project.

## 6. References

- P. F. Brown and al. 1993. The mathematics of statistical machine translation: parameter estimation. *Computational Linguistics*, 19:263–311.
- Tamas Varadi Jean Senellart, Pter Dienes. 2001. New generation systran translation system. In *MT Summit VIII*, Santiago de Compostela, Spain, September.
- Woosung Kim and Sanjeev Khudanpur. 2004. Lexical triggers and latent semantic analysis for cross-lingual language model adaptation. *ACM Transactions on Asian Language Information Processing (TALIP)*, 3(2):94–112.
- P. Koehn. 2004. Pharaoh: A beam search decoder for phrase-based statistical machine translation models. In *6th Conference Of The Association For Machine Translation In The Americas*, pages 115–224, Washington, DC, USA.
- Caroline Lavecchia, Kamel Smaili, and David Langlois. 2007a. Building parallel corpora from movies. In *Proceedings of The 5th International Workshop on Natural Language Processing and Cognitive Science*, Funchal, Madeira - Portugal, June.
- Caroline Lavecchia, Kamel Smaili, David Langlois, and J.P. Haton. 2007b. Using inter-lingual triggers for machine translation. In *Proceedings of the eighth conference in the annual series of INTERSPEECH*, Antwerp, Belgium, August.
- F. J. Och, C. Tillmann, and H. Ney. 1999. Improved alignment models for statistical machine translation. In *the joint conference of Empirical Methods in Natural Language Processing and Very Large Corpora*, pages 20–28, University of Maryland, College Park, MD.
- F.J. Och. 2002. *Statistical Machine Translation: From Single-Word models to Alignment Templates*. Ph.D. thesis, RWTH Aachen Department of Computer Science, Aachen, Germany.
- K. Papineni and al. 2001. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual of the Association for Computational linguistics*, pages 311–318, Philadelphia, USA.
- C. Tillmann and H. Ney. 1997. Word trigger and the EM algorithm. In *Proceedings of the Conference on Computational Natural Language Learning*, pages 117–124, Madrid, Spain.
- I. Zitouni, K. Smaïli, and J.-P. Haton. 2003. Statistical language modeling based on variable-length sequences. *Computer Speech and Language*, 17:27–41.